

Desarrollo de una aplicación ontológica para evaluar el mercado de trabajo español en Biblioteconomía y Documentación

Jose Antonio Moreiro*, Sonia Sánchez-Cuadrado**,
Jorge Morato**, Valentín Moreno**

Resumen: La tarea de evaluar la adecuación de las habilidades y competencias profesionales de los estudiantes de Biblioteconomía y Documentación es necesaria para el diseño de los planes de estudio de la especialidad. Sin embargo, esta labor conlleva un gran esfuerzo a los investigadores. Para facilitar el estudio se ha diseñado una metodología que simplifica su estudio periódico. Para validar la metodología se ha recopilado un corpus con los planes de estudio de la titulación y otro corpus formado por ofertas de empleo. A partir del corpus procedente de los planes de estudio se ha creado un tesoro que muestra la terminología del dominio. Con dicho tesoro se ha indexado automáticamente las ofertas de empleo. La comparación estadística entre los términos de las ofertas y los planes de estudio denota necesidades y carencias en la adecuación de los planes de estudio a las ofertas de trabajo. La metodología empleada permite presentar, de forma coherente y periódica la adecuación de planes de estudios a las demandas del mercado laboral. La comparación con estudios previos ha validado la idoneidad de la propuesta.

Palabras claves: Mercado de trabajo, Competencias, Biblioteconomía y Documentación, Formación universitaria, Actividad profesional, Tesoros.

An ontological method for evaluating the Spanish labour market in Library and Information Science

Abstract: *The assessment of how well the skills and professional competence of students of Library and Information Science meet current labour market demands is necessary for the design of appropriate curricula. Such work, however, requires a great deal of effort on the part of researchers. A methodology is here proposed that should facilitate the periodic studies that need to be made in this area. A corpus of Information and Library Science curricula and another reflecting the demands of the labour market were compiled. A thesaurus was developed from the former body of information in order to record the terminology used in this area. This was then used to index the demands made in offers of employment. The results highlight where our curricula fail to meet the*

* Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid. Correo-e: jamore@bib.uc3m.es.

** Departamento de Informática. Universidad Carlos III de Madrid. Correo-e: lssanchez, jorge, vmoreno@ie.inf.uc3m.es.

Recibido: 23-7-08; 2.^a versión: 22-9-08.

needs of the current labour market. The proposed method allows the coherence between curricula and labour market demands to be easily and periodically examined. The results were compared with those of previous studies in order to validate the proposed method.

Keywords: *Labour market, Library and Information Science, Competences, Skills, Professional activity, University training, Thesauri.*

1. Introducción

La Biblioteconomía y Documentación es un área en auge para el mercado de trabajo español. Conocer el tipo de perfil más demandado en las ofertas de empleo destinadas a la Biblioteconomía y Documentación tiene un indiscutible interés para la formación de estos profesionales.

La evaluación del mercado de trabajo español en el área de Biblioteconomía y Documentación ha sido analizada por entidades como la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) (2004) y la European Council for Information Associations (ECIA) (2004). Diferentes investigadores también han mostrado su interés sobre esta materia en diferentes publicaciones. Estos estudios sobre la adecuación de los diplomados y licenciados en Biblioteconomía y Documentación a los perfiles profesionales aplican principalmente dos enfoques metodológicos: 1) la elaboración de encuestas y entrevistas personales; 2) el estudio y la comparación manual de planes de estudio, ofertas y demandas de empleo.

1. En la elaboración de encuestas, es frecuente que el acceso a los encuestados se realice mediante la Web (Fujino, 2007; Moreiro et al., 2008), de forma telefónica (Tejada, 2003), y/o con entrevistas personales, entre otras. Estas estrategias tienen la ventaja de acceder a un tipo de información difícil de recoger por otras técnicas. Sin embargo, estas técnicas tienen el inconveniente de depender de la buena predisposición y objetividad del encuestado.
2. Para la comparación manual de planes de estudios, ofertas y demandas de empleo, Moreiro (1999) realizó su estudio utilizando los datos del Servicio de Orientación y Planificación Profesional (SOPP) de las universidades. El método aplicado por Moreiro se hace posible cuando se dispone de un amplio conocimiento sobre el campo de la Biblioteconomía y la Documentación. No obstante, este enfoque tiene el inconveniente de requerir gran dedicación y estudio.

Ambos enfoques suponen una alta laboriosidad dedicada a estos estudios. Por este motivo, tales investigaciones se acotan según las características y necesidades geográficas y/o temporales de las ofertas y demandas de empleo en

comparación con los planes de estudio. Algunos ejemplos de trabajos limitados por estas características son:

- Acotación geográfica o acotación según centro de estudios. Moreiro (1999) centró su evaluación inicialmente en la circunscripción de la Universidad Carlos III de Madrid, y posteriormente la amplió al territorio nacional (Moreiro 2001). También para los centros de estudios de la Universidad Complutense (Tejada, 2002) y la Universidad de Barcelona se han realizado análisis sobre la inserción laboral de los titulados.
- Acotación temporal. Los licenciados por la Universidad de Sao Paulo entre los años 2000 y 2004 también han sido objeto de estudio (Fujino, 2007). Otra evaluación de Moreiro et al. (2008) analiza el empleo en el sector de Biblioteconomía y Documentación entre los años 2000 y 2005.

2. Objetivos

El objetivo del presente trabajo es una propuesta metodológica que utiliza herramientas software y permite la automatización del estudio de la adecuación entre los planes de estudio y su correspondencia con las demandas del mercado de trabajo en el ámbito español de la Biblioteconomía y la Documentación. Esta propuesta pretende reducir el coste manual del estudio y eludir los sesgos indicados que proporcionan los estudios centrados en encuestas.

3. Metodología

La hipótesis de partida para la automatización del estudio de adecuación entre los planes de estudio y su correspondencia con las demandas del mercado de trabajo se basa en la creación y la comparación de dominios mediante sistemas de organización del conocimiento y procesos de indexación automática.

No se conocen trabajos previos que apliquen un enfoque automatizado en la identificación de perfiles profesionales y demandas de empleo. En el apartado siguiente se presentarán algunas iniciativas de otros ámbitos con las que se ha relacionado esta propuesta.

4. Antecedentes relacionados con la metodología

Sánchez-Cuadrado et al. (2007) propusieron un enfoque metodológico para mejorar los sistemas de transferencia de conocimiento entre los *callcenters* y los analistas de software de SAGE-SP. En esta propuesta similar, en concreto, se analizaban los textos de las llamadas, y se detectaban estadísticamente las propuestas de mejoras más demandadas por los usuarios del software de dicha empresa.

El proceso se realizaba mediante técnicas de procesamiento del lenguaje natural y la creación de diversos tesauros y ontologías (Sánchez-Cuadrado, 2007).

La identificación de términos procedentes de otro recurso en una ontología preexistente se utiliza en el presente estudio como base para identificar la frecuencia y las ausencias entre los términos del recurso y nuestra ontología ligera. Anteriormente un planteamiento similar se ha utilizado en la literatura para ampliar una ontología preexistente, ejemplos son FCA-Merge y Sensus.

FCA-Merge es un método para fusionar ontologías desde los conceptos específicos hacia los conceptos genéricos (Stumme y Madche, 2001). El proceso recibe como entrada dos ontologías que se desean fusionar y un conjunto de documentos cuyo contenido trata sobre el dominio de las ontologías (Gómez-Pérez, 2004). La fusión de las ontologías se consigue extrayendo instancias de los documentos que tienen una correspondencia con los conceptos de ambas ontologías. Si el concepto C1 de la ontología O1 tiene instancias en los mismos documentos que el concepto C2 de la ontología O2 entonces C1 y C2 son candidatos a ser considerados el mismo concepto. De este modo, la ontología fusionada se va «podando» para que el resultado contenga solo los términos existentes en los documentos, y los necesarios para unir esos términos. En conclusión, FCA-Merge es un método adecuado para ontologías con un nivel semántico similar a los tesauros.

Un diseño más próximo se realizó con la ontología Sensus, tanto en su propia construcción como para adaptar la ontología Sensus a las instancias de un nuevo dominio (Gómez-Pérez, 2004). Sensus¹ es una gran ontología desarrollada por USC/ISI durante la década de los años 90. Fue elaborada por la fusión de diferentes recursos como Penman Upper Model (Bateman, 1990), la ontología de alto nivel ONTOS (Carlson y Nirenburg, 1990), las categorías semánticas de LDOCE² y WordNet 1.4 (Fellbaum, 1998) y el diccionario the Harper-Collins Spanish-English Bilingual Dictionary. El proceso es el siguiente:

1. Identificar términos semilla (términos del dominio nuevo que deben existir en la ontología resultante) normalmente son específicos de conceptos existentes en la ontología Sensus.
2. Unir las instancias de un nuevo dominio a términos que existan en la ontología Sensus.
3. Añadir todos los términos desde el identificado de la ontología Sensus al término raíz de Sensus. El resultado es una subontología de la ontología original. Dicho de otro modo, se «podan» o eliminan aquellos nodos no existentes en el nuevo dominio y no necesarios para unir los conceptos identificados.
4. Por último, se añaden nuevos términos que no hayan aparecido, con la

¹ Ontology creation and use: Sensus <http://www.isi.edu/natural-language/resources/sensus.html>.

² Logman Dictionary of Contemporary English, publicado por primera vez en 1978. <http://www.ldoceonline.com>.

filosofía: si muchos términos del subárbol son relevantes todo el subárbol también lo es.

Según estos postulados, Rueda (2008) creó una aplicación software capaz de elaborar un tesauro especializado a partir de un tesauro preexistente y una colección de documentos representativos de la temática del tesauro inicial.

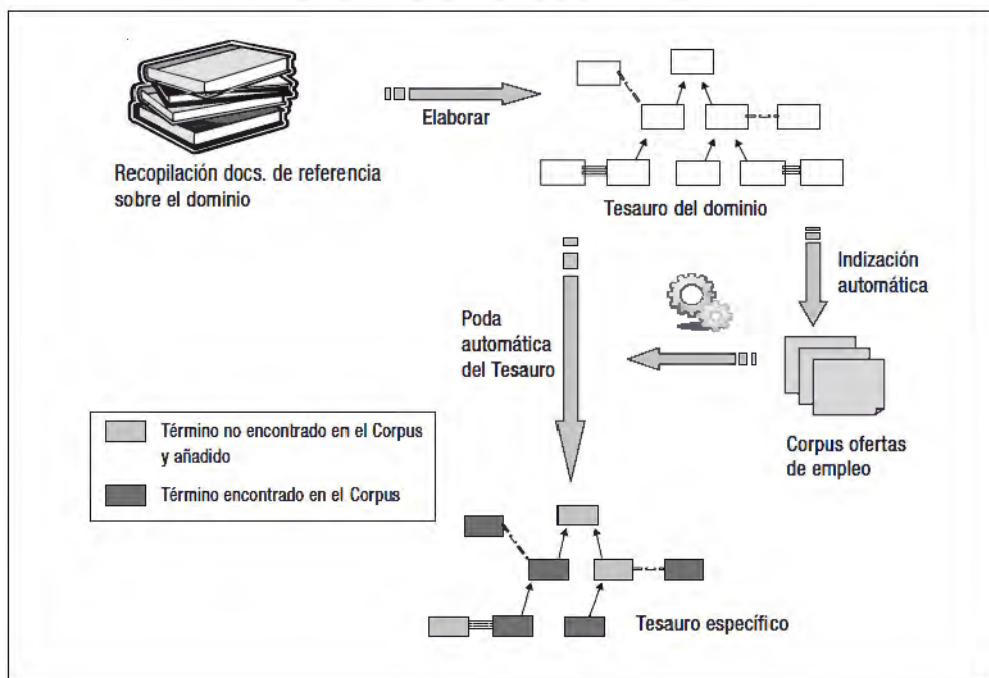
5. Método y recursos empleados

Los métodos y recursos empleados están orientados a la elaboración de representaciones de dominio mediante tesauros y que la especialización de los tesauros permita reconocer la información contenida en ofertas de empleo. Se parte de la necesidad de imbricar un conocimiento común que ha sido expresado de diferente modo a través del lenguaje y las palabras. La propuesta para unificar la variedad terminológica para un mismo referente se prevé solucionar con el tesauro mediante relaciones de generalización-especificación o sinonimia según proceda en cada caso.

La metodología para la elaboración del tesauro (Figura 1) se plantea de forma secuencial en las siguientes fases:

FIGURA 1

Metodología aplicada en el proyecto



1. Recopilación de recursos de información para la elaboración del tesaurus.
2. Elaboración manual del tesaurus sobre el dominio.
3. Recopilación de un corpus digital de ofertas de trabajo.
4. Indización automática del corpus digital de ofertas de trabajo con el tesaurus del dominio.
5. Elaboración del tesaurus específico.

6. Recopilación de recursos de información para la elaboración del tesaurus

Esta fase consiste en la selección de documentos que contengan terminología y conocimiento relevante para la elaboración de un tesaurus sobre los perfiles profesionales de los estudios de Biblioteconomía y Documentación. Forman parte del corpus los siguientes documentos:

- Libro blanco del título de grado en Información y Documentación (ANECA, 2004).
- Perfil Profesional de la Diplomatura de Biblioteconomía y Documentación: Análisis de la Demanda Gestionada en el COIE. Vicerrectorado de Estudiantes. Universidad Complutense de Madrid: 2004.
- Oferta de prácticas y empleo informe gestión SOPP 2007 de la Universidad Carlos III de Madrid [<http://www.fundacion.uc3m.es/Sopp/>].
- Plan de estudios del grado en Información y Documentación Univ. Carlos III de Madrid [www.uc3m.es].

7. Elaboración manual del tesaurus sobre el dominio

El tesaurus genérico se crea extrayendo los términos y relaciones del libro blanco de los estudios de Información y Documentación (ANECA, 2004) y los planes de formación en Biblioteconomía y Documentación y en Información y Documentación de la Universidad Carlos III. Se ha utilizado el editor Domain Reuser, para elaborar la estructura del tesaurus que permite su exportación e importación. Para dar mayor potencia al tesaurus creado se añadieron términos sinónimos y términos específicos procedentes de las ofertas de empleo.

8. Recopilación de un corpus digital de ofertas de trabajo

Se han recopilado 1087 ofertas de trabajo procedentes de la Web de Documentación (www.documentacion.com.es) entre el periodo de septiembre de 2005 y abril de 2008. Se ha escogido este recurso porque es el mayor portal especializado en el área de la Documentación en España. Se dedica sólo al ámbito es-

pañol, se organiza por años y meses y al no eliminar el histórico de peticiones se facilita el estudio temporal. Los autores de este trabajo son conscientes de que el estudio está limitado a un único portal, la Web Documentación [www.documentacion.com.es] (consultado julio 2008). Esta restricción puede proporcionar una visión limitada al no estar considerando otras ofertas de otros sitios web, que sin solicitarlo expresamente, definan un perfil de Biblioteconomía y Documentación.

9. Indización automática del corpus digital de ofertas de trabajo con el tesauro del dominio

Construido el tesauro en la primera fase, se indiza de forma automática el corpus digital de ofertas de trabajo. Mediante la indización se espera identificar los términos que aparecen en el corpus junto con su frecuencia. A efectos de este estudio, la principal característica de este sistema de indización automática es la capacidad de identificar términos compuestos (Sánchez-Cuadrado et al., 2007).

10. Elaboración del tesauro específico

La elaboración del tesauro específico se realiza mediante una herramienta software. El software empleado en este proyecto se ha desarrollado en lenguaje JAVA, bajo el mismo planteamiento implementado por Rueda (2008). La finalidad de la herramienta consiste en crear un subtesauro o tesauro parcial a partir de otro del dominio. El proceso detallado es el siguiente:

1. Importación y almacenamiento del tesauro. Se utiliza una base de datos referencial MySQL en respuesta a la necesidad de almacenar el tesauro de forma eficiente y rápida consulta. La importación del tesauro se realiza de forma automática a partir de un informe alfabético del tesauro.
2. Identificación de los descriptores que componen el vocabulario del subtesauro. El proceso consiste en identificar los términos del tesauro en los documentos. Para esta comparación es necesario aplicar reglas de lematización debido a la posible existencia de formas derivadas y flexionadas de los términos simples y compuestos que forman el tesauro. Además, se considera la posible existencia de palabras vacías entre los términos de los documentos, en el caso de que los descriptores del tesauro sean compuestos. Dado el uso incorrecto o ausencia de tildes en las palabras, se opta por no tener en cuenta los acentos por tratarse de un dominio específico. También, se procede de forma similar en la utilización de mayúsculas. Debido a que el orden de las palabras de un término compuesto no tiene un patrón fijo y dado que se trata de un dominio específico, se utiliza el criterio de proximidad de términos simples. De este

modo, dos términos compuestos como «Técnicas analíticas e instrumentales» y «Técnicas instrumentales y analíticas» se consideran el mismo descriptor.

3. Incorporación de los sinónimos al tesauro podado, ya aparezcan o no en el corpus. También se añaden las asociaciones entre términos que hayan aparecido en los documentos.
4. Eliminación o «poda» de los descriptores que o bien no hayan aparecido en los documentos o bien no sirvan para unir los descriptores jerárquicamente en *n* saltos (indicado por el usuario).
5. Las relaciones con términos no-descriptores y relacionados entre términos presentes en el subtesauro se importan directamente del tesauro genérico.
6. Establecimiento de las relaciones entre los términos del vocabulario. Una vez identificados los términos y para evitar «conceptos huérfanos» se unen los conceptos jerárquicamente por el camino más corto presente en el tesauro genérico. Esta acción implica añadir nuevos nodos que permitan unir estos términos. La aplicación requiere del usuario información sobre el número máximo de saltos, para evitar introducir un número excesivo de nodos en el caso de conceptos lejanos semánticamente.
7. Generación de un tesauro expresado en vista alfabética con una estructura de importación y exportación compatible con la aplicación Domain Reuser.

11. Resultados y discusión

El tesauro elaborado inicialmente tiene 812 términos y pertenecen a las siguientes familias: Búsqueda de empleo (demandas y ofertas de empleo), Condiciones laborales (contratos, salario, tipo de jornada), Empresas (área de actividad, sector de actividad, tamaño de empresa, tipo de empresa), Entidades e Instancias de Organismos, Formación (conocimientos, contenidos, formación reglada), Instrumentos (encuestas y entrevistas, foros, etc.), Lugares (topónimos), Perfil académico profesional, Perfil del solicitante (competencias, disponibilidad, edad, experiencia, habilidades, perfil ocupacional, perfil prácticas), Personas y Procesos (acreditación, gestión de contenido, etc.).

Del proceso de indización de las ofertas de empleo, se identificó un total de 269 términos, es decir un 33% de coincidencia con los términos del tesauro. Se añadieron 27 términos para unir los distintos subárboles, de modo que el resultado del tesauro específico queda con un total de 296 términos, de los cuales 76 eran no-descriptores (sinónimos). Los siguientes descriptores han sido identificados en el tesauro específico, si bien no se trata de un listado completo.

- Áreas de actividad: Administración, Agencia de comunicación, Alimentación, Bancos, Comercio, Construcción, Consultoras, Editoriales, Empresas informáticas, Empresas de contenidos, Financiera, Hostelería, Industria, Institu-

ción Educativa, Investigación, Legales, Prensa, Recursos Humanos, Salud, Seguros, Servicios, Transportes.

- Conocimientos demandados: Archivística, Biblioteconomía, Conocimiento de idiomas, Documentación e información, Empresariales, Informática, Lingüística, Tratamiento de imágenes.
- Administración de bibliotecas: Análisis de fuentes documentales, Atención al usuario de la biblioteca, Catalogación del fondo bibliográfico, Gestión bibliotecaria, Inventario, Organización de bibliotecas.
- Tipos de bibliotecas que más demandan: Bibliotecas en centros de enseñanza, Bibliotecas especializadas, Bibliotecas musicales, Bibliotecas públicas, Bibliotecas escolares, Bibliotecas universitarias.
- Competencias demandadas: Capacidad de comunicación, Capacidad de trabajo, Competencias específicas de la titulación, Competencias transversales (Instrumentales, Capacidad de análisis y síntesis, Resolución de problemas, Toma de decisiones), Personales (Capacidad de trabajar en equipo, Habilidades en las relaciones interpersonales) y Sistémicas (Creatividad, Iniciativa y espíritu emprendedor, Liderazgo), Responsabilidad, Competencias en formación, Competencias en técnicas comerciales.
- Idiomas: alemán, castellano, catalán, francés, inglés, portugués.
- Informática: Lenguajes de programación (Java, Java Script, Perl, PHP), Ofimática, Sistemas operativos (Linux, Windows), Web (CMS, Joomla, Diseño Web, Lenguajes de etiquetado (XML, RDF, OWL, HTML), Posicionamiento Web, Sindicación de contenidos, Gestión de bases de datos (Access, MySQL).
- Representación y recuperación de información (Catalogación descriptiva, Clasificación indización, Lenguajes documentales, Recuperación de información, Resúmenes), Sistemas de organización del conocimiento (Taxonomías, Thesaurus, Vocabularios de metadatos), Tecnologías de la información (Pruebas de aplicaciones informáticas, Postventa), Tratamiento de la información (Búsquedas en Internet, Escaneado, Gestión de bases de datos, Tratamiento informático de la información, Elaboración de informes sobre la competencia), Gestión de la información (Archivo documental, Búsqueda de documentación, Mantenimiento de archivos, Realización de búsquedas de información).
- Perfil académico (Diplomado, Licenciado, Postgraduado), Perfil ocupacional (Administración de bibliotecas, Comercial, Formación de personal, Gestión de la información, Tecnologías de la información, Tratamiento de la información, Asistente de proyectos), Perfil profesional (Ayudante de documentalista, Bibliotecarios escolares, Control de calidad de fondo antiguo, Digitalización de fondo antiguo, Documentalista, Gestor de contenido, Grabador de datos, Linkbuilder, Personal de archivos, Webmaster).

Para poder evaluar la calidad de los resultados y analizar la capacidad de la aplicación mediante el reconocimiento terminológico y el cálculo Tf, se selecciona al azar un subconjunto de 200 demandas del corpus de ofertas de empleo.

Este subconjunto reducido permite ser indizado y revisado. Los resultados son comparados con el informe del COIE de la UCM en 2004 y el estudio de Moreiro (2008) referente a las áreas de actividad, las aptitudes y las tareas que se demandan en las ofertas laborales para profesionales de la Biblioteconomía y la Documentación.

En la Tabla I se muestran los resultados de frecuencia obtenidos para los tipos de áreas de actividad. Haciendo uso de las propiedades de los tesauros, se debe tener en cuenta que los valores de ocurrencias, en este caso, agrupan al descriptor y al no-descriptor (sinónimo), pero no a sus específicos.

Áreas de actividad

TABLA I

Resultado de frecuencia de los descriptores de áreas de actividad

Término	Frecuencia %
Prensa	20,0
Institución educativa	18,2
Bancos	12,7
Editoriales	12,7
Comercio	9,1
Salud	7,3
Recursos humanos	7,3
Industria	7,3
Consultoría	3,6
Financiero	1,8

Si se comparan los resultados de la Tabla I con los resultados del informe del COIE de la UCM en 2004, se aprecian algunas diferencias, si bien se observa que los grupos principales coinciden (marcados en gris en la Tabla I).

- M. de comunicación y agencias..... 32,46%
- Consultoría, asesoría y auditoría 12,69%
- Telecomunicaciones..... 11,57%
- Servicios a empresas 9,70%
- Servicios a la comunidad..... 7,09%
- Editorial y artes gráficas..... 5,41%
- Informática 4,66%
- Industria..... 3,17%
- Banca y seguros 2,99%

Según el estudio de Moreiro et al. (2008) realizado con encuestas, los entrevistados respondieron con la siguiente información:

• Administración Pública.....	43,4%
• Consultoría general.....	9,3%
• Medios de comunicación	7,6%
• Sector informático y telecom.....	6,7%
• Fundaciones y ONG	6,1%
• Consultoría documental.....	4,6%
• Editoriales	1,5%

Algunas de las diferencias entre estos estudios y los resultados de la aplicación se deben realmente al origen de las demandas utilizadas en los experimentos, ya que la Administración Pública o las ONG no suelen publicar sus demandas en portales de Internet para documentalistas.

Aptitudes

Los resultados del proceso automático centrado en aptitudes se muestran en la Tabla II.

A diferencia del informe de Moreiro et al. (2008) donde las 21 aptitudes están entre el 3,5% y el 6%, se ha detectado una significativa reducción de las aptitudes requeridas y un aumento considerable en la iniciativa y el espíritu emprendedor así como en la capacidad de trabajar en equipo.

TABLA II

Resultados de frecuencia de los descriptores de aptitudes

Término	Frecuencia %
Iniciativa y espíritu emprendedor	41,1
Capacidad de trabajar en equipo	27,4
Capacidad de trabajo	12,6
Responsabilidad	10,5
Creatividad	8,4

Conocimientos y tareas

El análisis de las tareas más demandadas en las ofertas de trabajo reflejó con claridad un perfil profesional vinculado al área de informática. Los autores lo encontraron notablemente significativo y se ha reflejado por separado en otra tabla (Tabla III).

TABLA III

Resultados de frecuencia de los descriptores de tareas de informática

Término	Frecuencia %	Descriptor genérico
Web	29,6	Informática
Word	8,7	Procesador de textos
Java script	7,5	Lenguajes de programación
Seo	7,0	Posicionamiento web
Usabilidad	5,2	Diseño web
Xml	5,0	Lenguajes de etiquetado
Power point	4,2	Presentaciones
Php	4,0	Lenguajes de programación
Diseño web	3,7	Web
Accesibilidad	3,0	Diseño web
Windows	2,7	Sistemas operativos
Posicionamiento web	2,7	Web
Webmaster	2,2	Perfil profesional
Java	2,0	Lenguajes de programación
Presentaciones	2,0	Ofimática
Programas de mensajería	1,7	Ofimática
Sql	1,7	Bases de datos
Ocr	1,5	Escaneado
Sistemas operativos	1,0	Informática
Portales de Internet	1,0	Contenidos
Linux	0,7	Sistemas operativos
Bases de datos documentales	0,7	Bases de datos
Rdf	0,5	Lenguajes de etiquetado
Perl	0,2	Lenguajes de programación
Procesador de textos	0,2	Ofimática
Ayuda on-line	0,2	Tratamiento de la información
Arquitectura web	0,2	Diseño web
Mantenimiento de bases de datos	0,2	Gestión de bases de datos

En el informe del COIE de la UCM, en 2004, también se señaló como un aspecto destacable el componente informático en los currículos en Biblioteconomía y Documentación, así los conocimientos más solicitados eran:

- Ofimática 15,95 %
- Access 12,77 %
- Otras bases de datos 12,77 %
- Word 9,57 %
- Excel 7,45 %
- Lenguajes de programación 7,45 %
- Sistemas operativos 7,45 %
- Tratamiento de imágenes 6,38 %

Al contrastar los datos del COIE de 2004 con los obtenidos en el estudio hay que destacar el aumento de la importancia de la informática en los últimos tres años. Pero además, como se puede observar en la Tabla III, los conocimientos que se exigen, superan el nivel de usuario de informática.

A continuación, se ha realizado un análisis similar, pero con los conocimientos más demandados en Documentación, excluyendo los ya señalados en la tabla de resultados de tareas de Informática.

TABLA IV
Resultados de frecuencia de los descriptores de tareas

Término	Frecuencia %	Descriptor genérico
Documentalista	56,0	Perfil profesional
Bibliotecas	12,5	Entidades
Conocimiento de idiomas	5,2	Conocimientos
Publicidad y marketing	4,3	Comercial
Proveedores	3,9	Comercial
Entrevistas	2,0	Instrumentos
Escaneado	2,0	Tratamiento de la información
Lenguajes documentales	1,6	Representación y recuperación de la información
Resumen	1,6	Representación y recuperación de la información
Thesaurus	1,6	Sistemas de organización del conocimiento
Inventario	1,3	Administración de bibliotecas
Video	1,3	Formato de la documentación
Investigación comercial	1,0	Comercial
Mantenimiento de archivos	1,0	Gestión de la información

TABLA IV (continuación)

Término	Frecuencia %	Descriptor genérico
Biblioteconomía	0,7	Conocimientos
Asistente de proyectos	0,7	Perfil profesional
Vocabularios de metadatos	0,7	Sistemas de organización del conocimiento
Soporte a clientes	0,7	Soporte técnico
Tejuelos	0,3	Biblioteconomía
Grabaciones sonoras	0,3	Formato de la documentación
Archivo documental	0,3	Gestión de la información
Bibliotecarios escolares	0,3	Perfil profesional
Dublin core	0,3	Vocabularios de metadatos

Moreiro et al. (2008) en su informe obtuvo los siguientes resultados para las tareas más demandadas:

- Descripción de materiales/catalogación 8%
- Elaboración, implementación y mantenimiento de bases de datos 6,9%
- Organización del archivo..... 6,8%
- Prestación de servicios de búsqueda de información y acceso a BD..... 6,4%
- Indización y clasificación..... 6,3%
- Préstamo de materiales 5,7%
- Formación de usuarios..... 4,6%
- Elaboración y mantenimiento de portales y páginas WEB .. 4,1%
- Orientación bibliográfica..... 3,6%
- Elaboración y actualización de lenguajes documentales 3,4%
- Adquisición documental..... 3,2%
- Planificación de los espacios, equipamientos, y señalización 3,2%

La comparación es complicada, debido a diferencias en el vocabulario y la granularidad entre la descripción de la oferta y las clasificaciones realizadas mediante encuestas. De este modo, términos como «descripción de materiales», «catalogación» e «indización» del estudio de Moreiro (2008) pueden estar expresadas en las ofertas bajo términos como «documentalista», «bibliotecas», «lenguajes documentales», «thesaurus» o «biblioteconomía». «Orientación bibliográfica» y «formación de usuarios» podrían estar asociados a los términos de las ofertas que especifican «soporte a clientes» o «bibliotecas», etc. Salvadas estas dificultades terminológicas se ven asociaciones entre ambas tablas. Aunque existe un perfil que significativa-

mente no aparece en estudios anteriores como la formación «comercial» (presente en términos como «publicidad y marketing», «proveedores», «soporte a clientes» o «investigación comercial», y en otros de la tabla de informática como «herramientas SEO» o «posicionamiento»).

12. Carencias y dificultades encontradas

De la comparativa entre los términos obtenidos de las 200 ofertas de trabajo y los términos del tesaurus específico se detecta un conjunto de descriptores que no han sido identificados en las ofertas en el proceso de indización. Un análisis sobre los descriptores del tesaurus no localizados en el corpus permite concluir que:

- El tesaurus no posee la cobertura deseada. Algunos ejemplos de las carencias en el tesaurus se localizan en el modo de designar lenguas como el «Catalán» y el «Gallego». En los documentos se han encontrado referencias a «Conocimiento de lenguas cooficiales del Estado». Otro caso de ejemplo se ha encontrado para señalar un perfil de «Castellano» que era referenciado por «hablante de español». Este tipo de carencias pueden ser corregidas en futuras versiones del tesaurus mediante la incorporación del término y la relación correspondiente.
- Los términos muy específicos como «Agència per a la Qualitat del Sistema Universitari» o «Aix-Marseille II» tienden a no aparecer en las ofertas de empleo. No obstante, los autores han conservado estos términos en el tesaurus, porque, en próximas versiones de esta propuesta, se espera analizar también los Curriculum Vitae de los solicitantes y estudiar su idoneidad con la oferta de trabajo.
- Los términos extraídos del libro blanco para el tesaurus de origen pertenecen a contextos distintos y no siempre se reflejan en las ofertas. «Razonamiento Crítico en el Análisis», «Integración en Equipos Multidisciplinares» o «Aprendizaje Autónomo». Estos términos, de forma modificada, están presentes en ocasiones en las ofertas, por lo que dado el caso, puede ser resuelto con relaciones de sinonimia.

13. Conclusiones y trabajos futuros

De la presente investigación se han extraído conclusiones relativas a varios aspectos: el tesaurus, las herramientas y la metodología.

En cuanto a la metodología, destacar que se ha propuesto un método automatizado que permite estudiar la adecuación de los planes de estudio a las ofertas laborales para profesionales de la biblioteconomía y la documentación. La metodología sería completa si se pudiera recoger información sobre cuales son las demandas que han sido satisfechas por el mercado. También podría ser com-

plementado con las percepciones del usuario, por ejemplo la valoración sobre la «formación recibida» (Moreiro, 2008 et al.).

El resultado de la aplicación software permite realizar estudios reduciendo el coste manual de la comparativa, así como eludir los sesgos temporales y de localización de otros estudios. Este enfoque evita realizar encuestas y acotar los resultados a un conjunto limitado de consultas. Asimismo permite acceder a perfiles que normalmente no se asocian a los estudios de Biblioteconomía y Documentación y proporciona mayor agilidad en la elaboración de nuevos informes, etcétera. En cuanto a las conclusiones, aumenta la credibilidad de los resultados.

El tesoro inicial del dominio tiene cerca de mil términos. Se ha implementado una aplicación capaz de generar un tesoro específico a partir de otro y por tanto el estudio se puede realizar tantas veces como se quiera aplicar a diferentes corpus de ofertas de trabajo. Una ventaja del planteamiento basado en tesoros es la capacidad del tesoro para agrupar los conceptos más específicos de la jerarquía, con lo que los informes pueden graduarse según distintos niveles de abstracción.

Entre los resultados cabe mencionar el gran peso que tiene la informática en las ofertas de perfiles documentalistas, sobrepasando los descriptores del área a los más específicos de la documentación, es decir en las demandas de profesionales se solicitan perfiles con conocimientos avanzados de aspectos informáticos. En el estudio se han observado ausencias en estudios anteriores de la vertiente comercial de la profesión.

Se seguirá trabajando en la mejora de las herramientas software y se espera que proporcionen un potencial equivalente a los trabajos referenciados. Sin embargo, los resultados obtenidos para el presente estudio, permiten validar el potencial de esta propuesta automatizada. Se ha constatado la idoneidad de la respuesta, su menor coste y el gran potencial que supone este enfoque para el futuro. La propuesta tiene un alto grado de reutilización: bien en estudios idénticos, pero efectuados en distintos años o ámbitos, donde la reutilización del tesoro apenas tiene coste de adaptación; bien en otros entornos en los que se quiera comparar, por cualquier motivo, dos dominios.

Por último, los autores del artículo aspiran a la automatización total del proceso para realizar este tipo de estudio de forma periódica y poder elaborar informes de forma automatizada que presenten la comparación y disminución de los costes que conlleva el estudio, así como reducir posibles sesgos y acotaciones. Por otra parte, está en desarrollo la elaboración de una comparación entre la situación de Brasil y de España. Se está confeccionando el tesoro y las herramientas lingüísticas de modo bilingüe.

14. Agradecimientos

Este estudio ha sido financiado por los programas de formación de la Fundación Carolina y el Programa Hispano-Brasileño de Cooperación Interuniversitaria del Ministerio de Ciencia e Innovación.

Referencias

- ANECA (2004): Libro blanco del título de grado en información y documentación. Madrid, España: Agencia Nacional de Evaluación de la Calidad y Acreditación, p. 238.
- Bateman J. A. (1990): Upper modeling: organizing knowledge for natural language processing. In *5th. International Workshop on Natural Language Generation*, Pittsburgh, PA, June.
- Borrego, A., Comalat, M. y Estivill, A. (2004): Inserció laboral dels titulats en Biblioteconomia i Documentació per la Universitat de Barcelona. En: *BiD, textos universitaris de biblioteconomia i documentació*, n.º 12, juny. Disponible en: http://www.ub.es/bid/consulta_articulos.php?fichero=12borreg.htm (consultado el 2 de julio 2008).
- Carlson, L. y Nirenburg, S. (1990): World modeling for NLP. Technical Report CMU-CMT-90-121, Carnegie Mellon University, Center for Machine Translation, Pittsburgh, PA.
- ECIA (2004): Euroreferencial en Información y Documentación. Competencias y aptitudes de los profesionales europeos de información y documentación. 2 ed. Madrid, España: SEDIC, European Council of Information Associations, p. 127. <http://www.certidoc.net> (consultado el 2 de julio de 2008).
- Fellbaum, C. (ed.) (1998): WordNet: An On-line Lexical Database and Some of its Applications. MIT Press, Cambridge, MA.
- Fujino, A., Lara, M., Amaro, R., Vergueiro, W., Tejada Artigas, C. y Moreira, J. A. (2007): Perfil profissional e competências de informação em Biblioteconomia e Documentação: pesquisa com egressos da ECA/USP. In: XXII CBBBD, Brasília. Anais XXII CBBBD. Brasília: FEBAB.
- Gómez-Pérez, A., Fernández-López, M. y Corcho, O. (2004): Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web, London, Springer, p. 403. ISBN 1852335513.
- Moreiro González, J. A. (1999): Ajuste de los licenciados en documentación a la oferta de trabajo. Observaciones hechas en la Universidad Carlos III de Madrid. En: *Revista General de Información y Documentación*, vol. 9, n.º 2, pp. 13-28.
- Moreiro González, J. A. (2001): «Figures on employability of Spanish Library and Information Science graduates». *Libri*, vol. 51, pp. 27-37.
- Moreiro González, J. A., Azcárate-Aguilar-Amat, P., Marzal-García-Quismondo, M. A., Tejada Artigas, C. M. y Vergueiro, W. (2008): Desarrollo profesional y opinión sobre la formación recibida de los titulados universitarios en información y documentación de las universidades públicas de Madrid (2000-2005). *El Profesional de la Información*, vol. 17, n.º 3, pp. 261-272.
- Rueda, A. (2008): Construcción de ontologías a partir de la especialización de fuentes estructuradas. Proyecto Fin de Carrera dirigido por J. Morato y V. Moreno. Madrid: Universidad Carlos III.
- Sánchez-Cuadrado, S. (2007): Definición de una metodología para la construcción automatizada de sistemas de organización del conocimiento. Tesis doctoral. Leganés: Universidad Carlos III.
- Sánchez-Cuadrado, S., Morato, J., Moreira, J. A. y Marrero, M. (2007): Definición de una metodología para la construcción de sistemas de organización del conocimiento a partir de un corpus documental en lenguaje natural. *Revista de la SEPLN*, vol. 39,

pp. 213-220. XXIII Congreso de la SEPLN. Universidad de Sevilla 10, 11, 12 de septiembre.

Stumme, G. y Madche, A. (2001): FCA-Merge: Bottom-up merging of ontologies. In 7th Intl. Conf. on Artificial Intelligence (IJCAI '01), pp. 225-230, Seattle, WA. <http://citeseer.ist.psu.edu/stumme01fcamerge.html>.

Tejada Artigas, C. (2002): Adecuación de los planes de estudio de la Diplomatura en Biblioteconomía y Documentación a las demandas del mercado de trabajo. El caso concreto del nuevo plan de estudios de la Escuela Universitaria de Biblioteconomía y Documentación de la Universidad Complutense de Madrid. Tesis doctoral. Madrid: Universidad Carlos III.

Tejada Artigas, C. y Moreiro González, J. A. (2003): Mercado de trabajo en Biblioteconomía y Documentación. Estudios sobre la inserción laboral de los titulados universitarios. En *El profesional de la información*, vol. 12, n.º 1, enero-febrero, pp. 4-9.